Marker development for the study at micro- and macro-evolutionary time scales in neotropical Palms

Marylaure de la Harpe, Oriane Loiseau, Jaqueline Hess, Nicolas Salamin, Christian Lexer, **Margot Paris**

UNI FR
UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

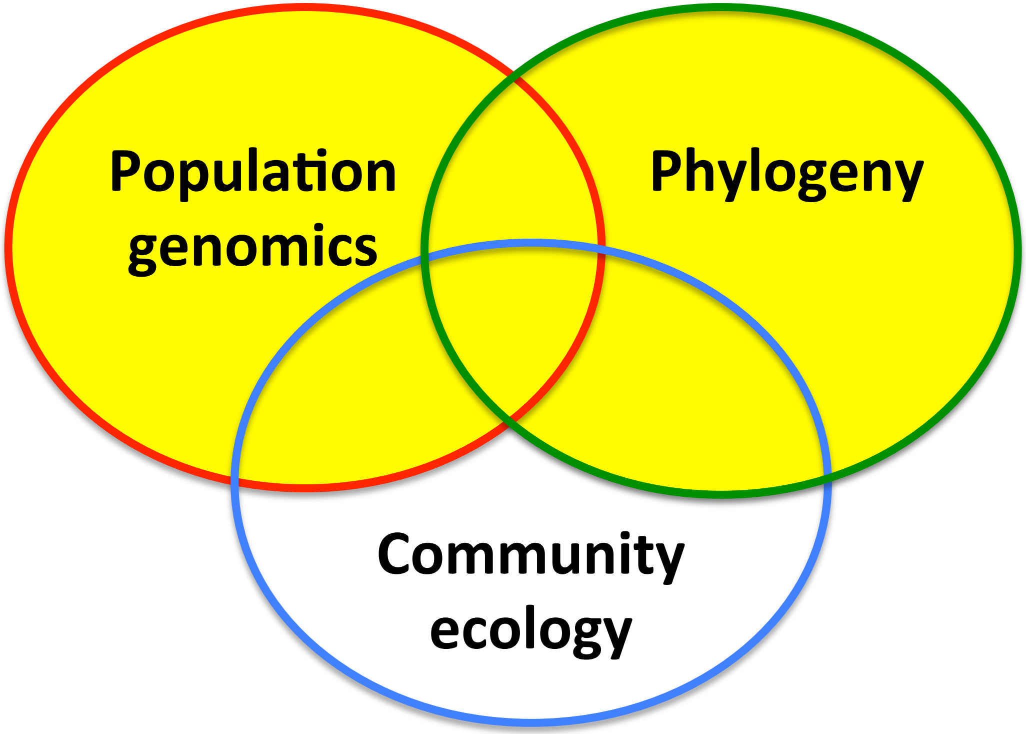
universität wien

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

(picture: Oriane Loiseau)

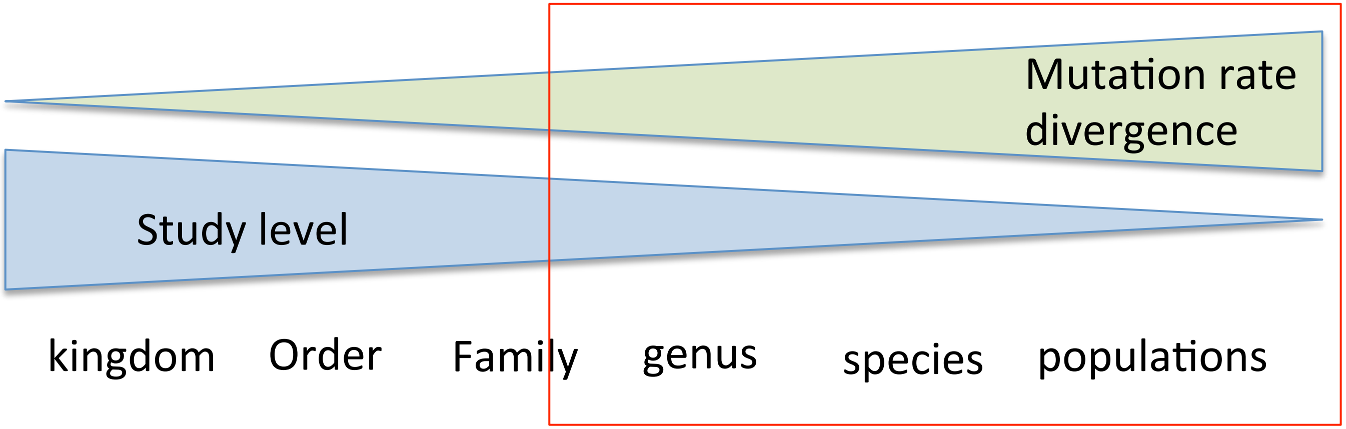
POPCORN, a multidisciplinary project

Using Population Genomics, Phylogenetics and Community Ecology to understand Radiations in Neotropical mountains



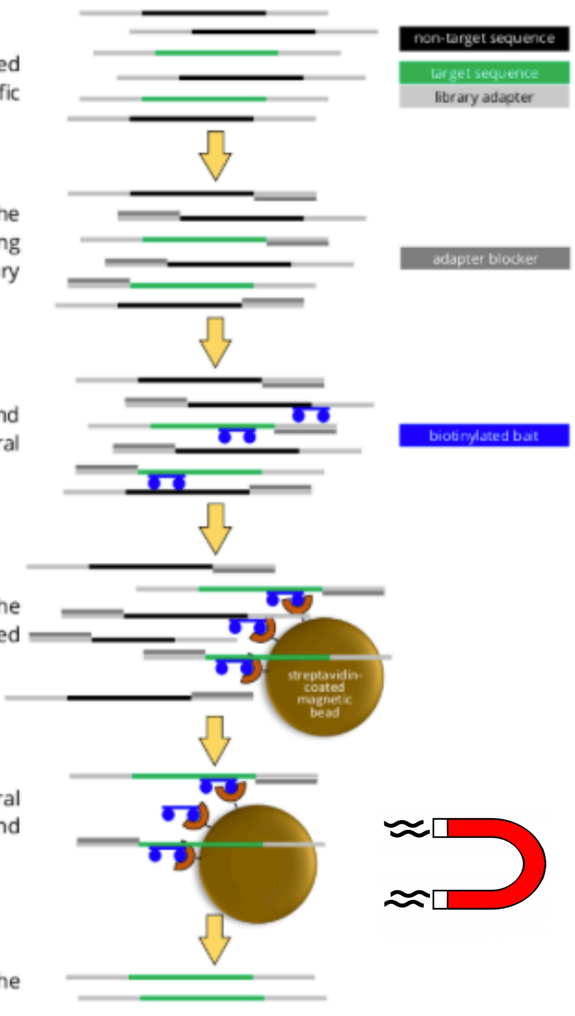
Ideal markers

- Many markers widespread along the genome
- Low cost in order to genotypes thousands of samples
- Evolution rate suitable for both macro and micro evolution studies



- Long sequences (>600 bp) for phylogeny and selection tests
- Include candidate genes for adaptation and “neutral” non-genic markers
- Include markers already used for phylogeny in palms
- Can be applied to low quantity and quality DNA from herbarium specimens

Target capture sequencing



1) DNA sequencing library is heat-denatured in the presence of adapter-specific blocking oligonucleotides

2) Library and blockers are dropped to the hybridization temperature, allowing blockers to hybridize to the library adapters

3) Biotinylated RNA baits are introduced and allowed to hybridize to targets for several hours

4) Bait-target hybrids are pulled out of the solution with streptavidin-coated magnetic beads

5) Beads are stringently washed several times to remove non-hybridized and nonspecifically-hybridized molecules

6) Captured DNA library is released from the beads and amplified

Very flexible as we can choose the targets:

- -number
- -nature (genes, non-genic regions)
- -candidate genes/regions
- -location in genome
- -length (in bp)
- -…

MYcroarray
THE OLIGO LIBRARY COMPANY

MYbaits®

Customized target enrichment kits for next-gen sequencing

arbor biosciences

http://www.arborbiosci.com/products/custom-target-capture/

Oil palm genome: very useful

- Oil palm genome is the closest reference genome

too divergent for proper capture design ??

especially because we are not interested in conserved regions

Building Geonoma reference sequences

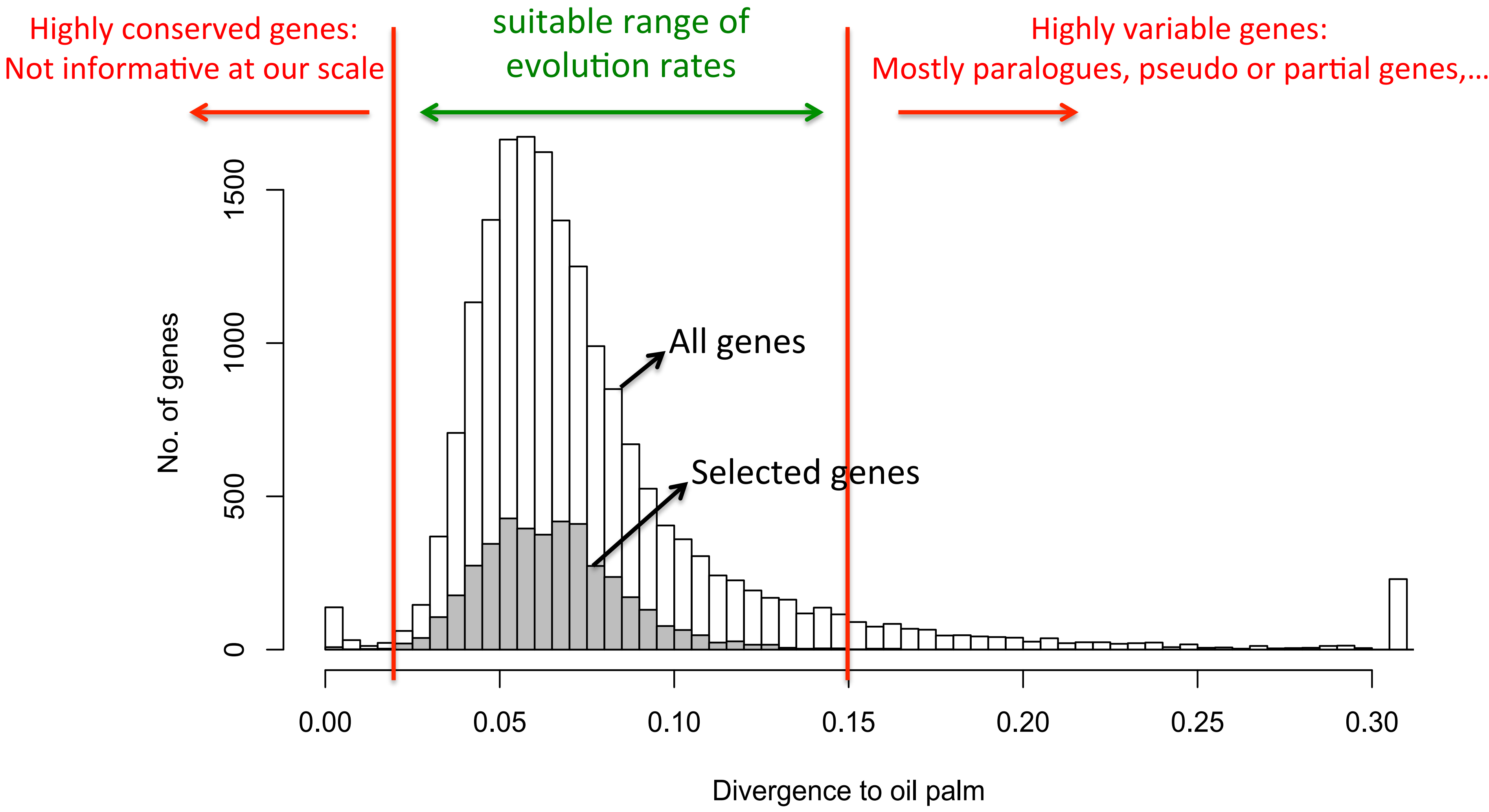
- Whole genome sequencing of the species *G. undata* (27x coverage, Illumina PE150bp)
- Reference assisted reconstruction of the *G. undata* genome
  - ➢ 94% of the genes recovered (UTRs + exons + introns)
  - ➢ Low recovery of the inter-genic regions (repeats, too divergent to the oil palm,…)

Criteria for the selection of 4’051 genes

- Broad range of rates of molecular evolution

Criteria for the selection of 4’051 genes

- Divergence to oil palm used as proxy for rate of molecular evolution



Highly conserved genes:
Not informative at our scale

suitable range of
evolution rates

Highly variable genes:
Mostly paralogues, pseudo or partial genes,…

Criteria for the selection of 4’051 genes

- Broad range of rates of molecular evolution
- Mostly single copy genes (using coverage and He info)
- Average size of 1’300 bp
- Interesting functions: pathogenesis; flowering; response to UV, light; floral scents;…
- 8 genes previously used for phylogeny + 141 Heyduk et al. (2015) genes
- Even distribution in the genome (around 160Kb on average between 2 target genes)

Additional 133 non-genic regions

- 5 to 15 per chromosomes
- 800 bp length in average
- As far as possible from genes

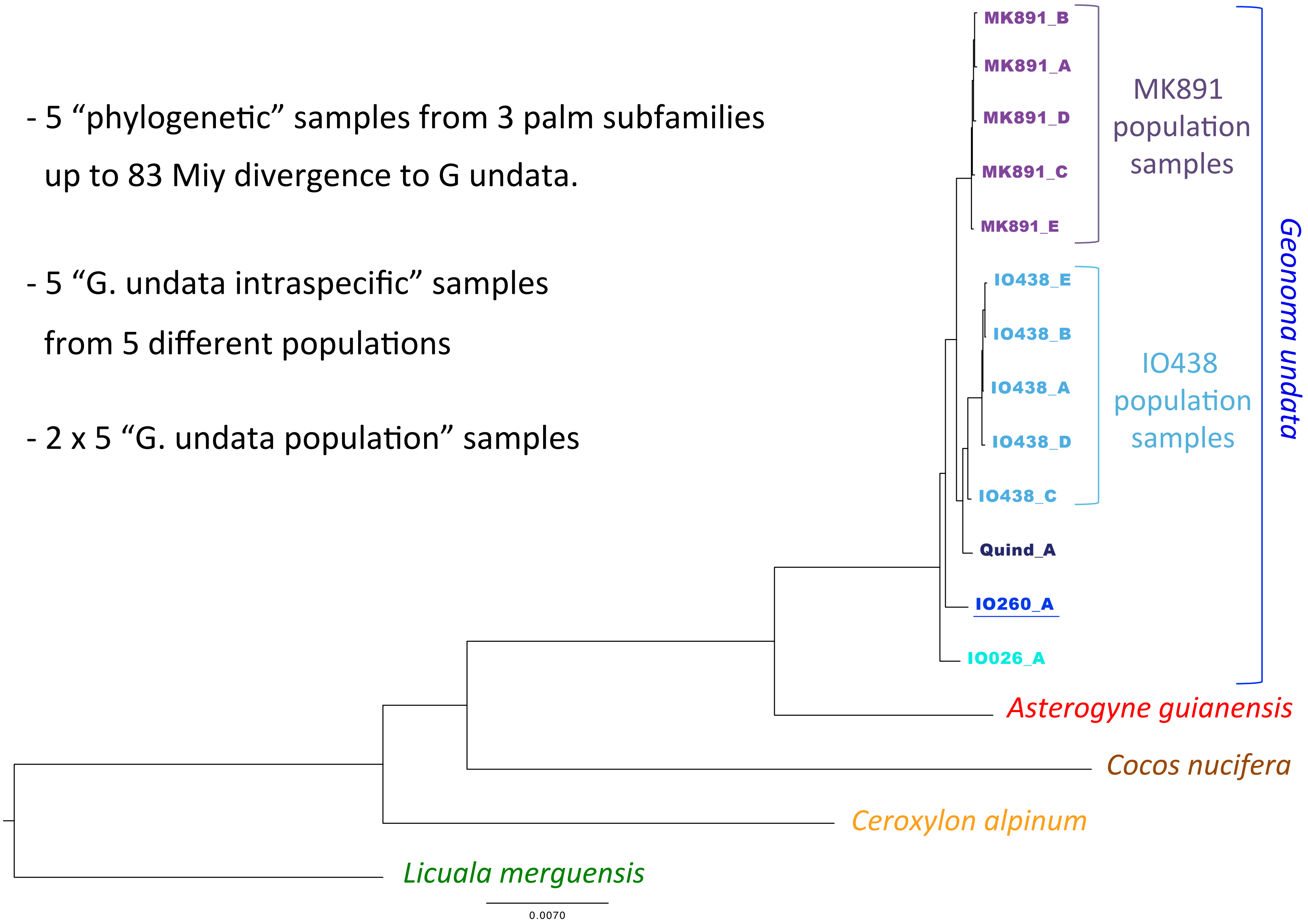
→ **4’770’883 bp in total**

Sampling for kit evaluation

- 5 “phylogenetic” samples from 3 palm subfamilies up to 83 Miy divergence to G undata.

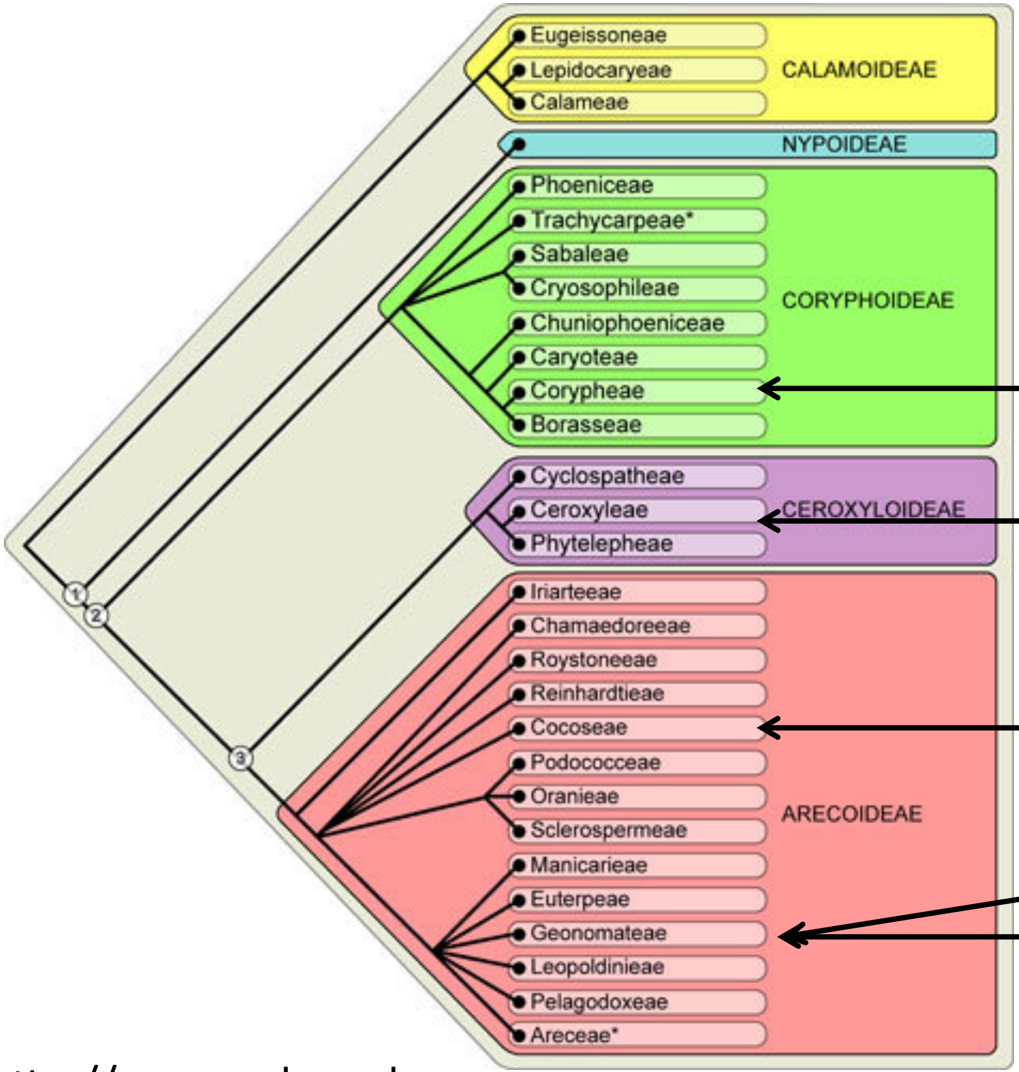
- 5 “G. undata intraspecific” samples from 5 different populations

- 2 x 5 “G. undata population” samples



Sampling for kit evaluation : phylogeny samples

- **3 palm subfamilies represented**
- **Up to 80 Miy divergence to *G. undata***



Licuala merguensis

Ceroxylon alpinum

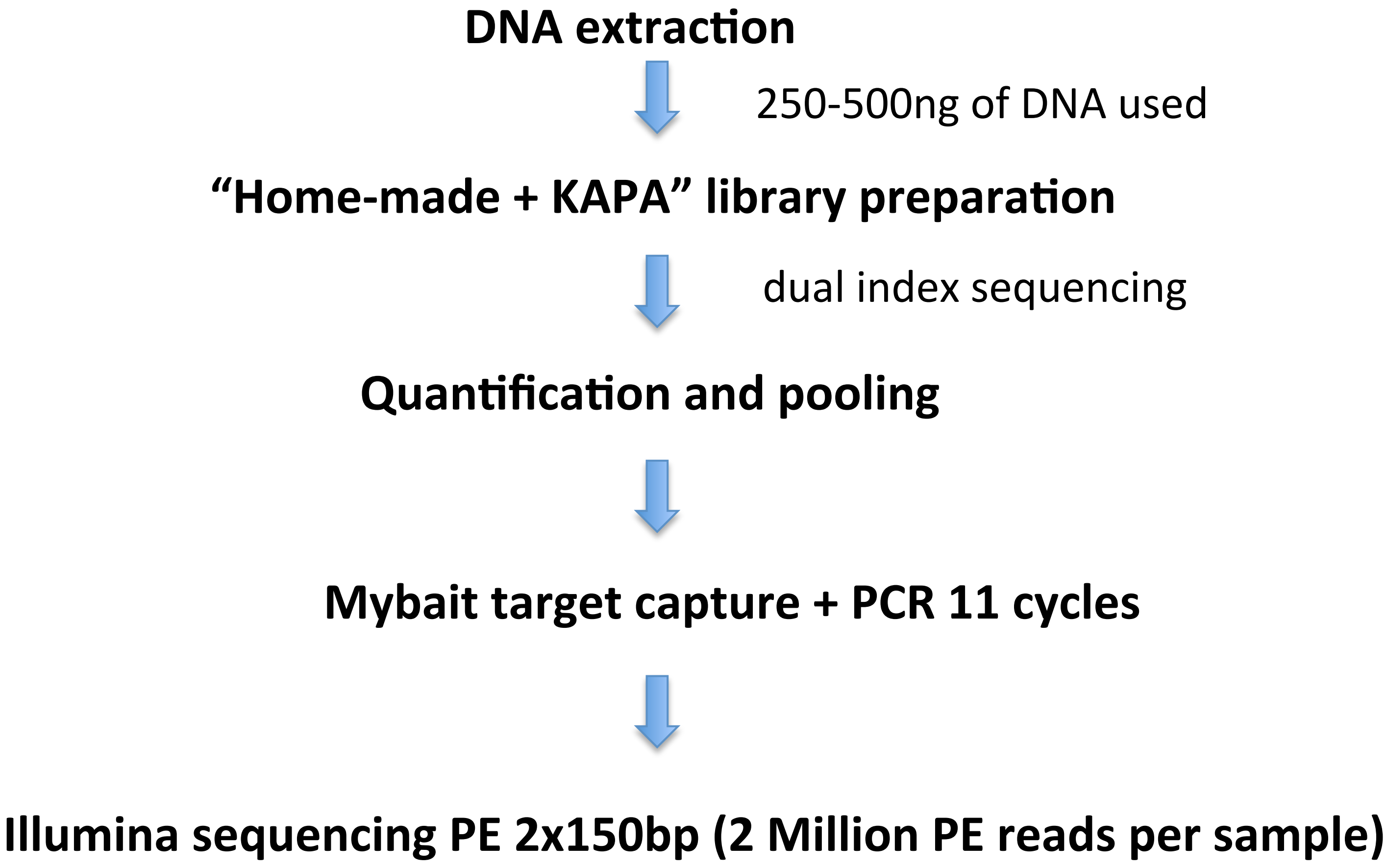
Cocos nucifera

Asterogyne guianensis

Geonoma undata

http://www.palmweb.org

Protocole



DNA extraction

250-500ng of DNA used

"Home-made + KAPA" library preparation

dual index sequencing

Quantification and pooling

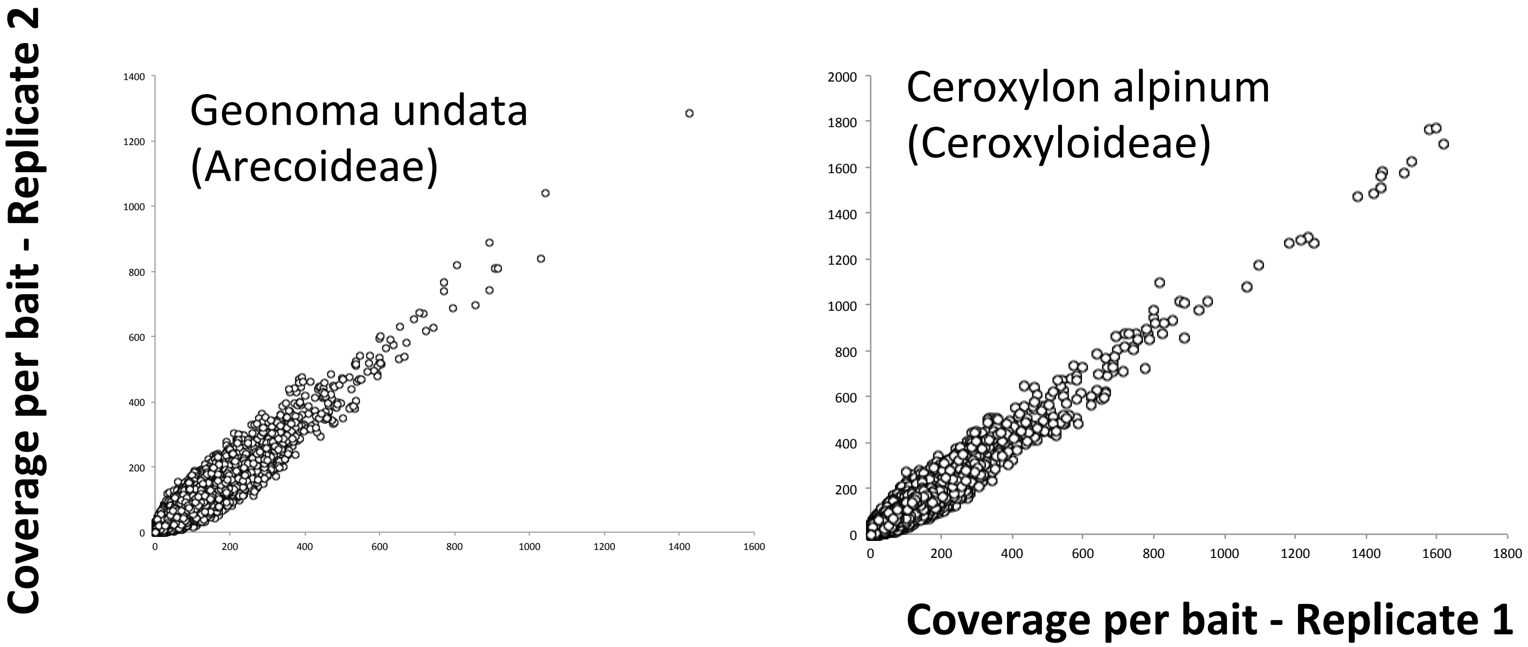
Mybait target capture + PCR 11 cycles

Illumina sequencing PE 2x150bp (2 Million PE reads per sample)

→ Total cost per sample = 80 $

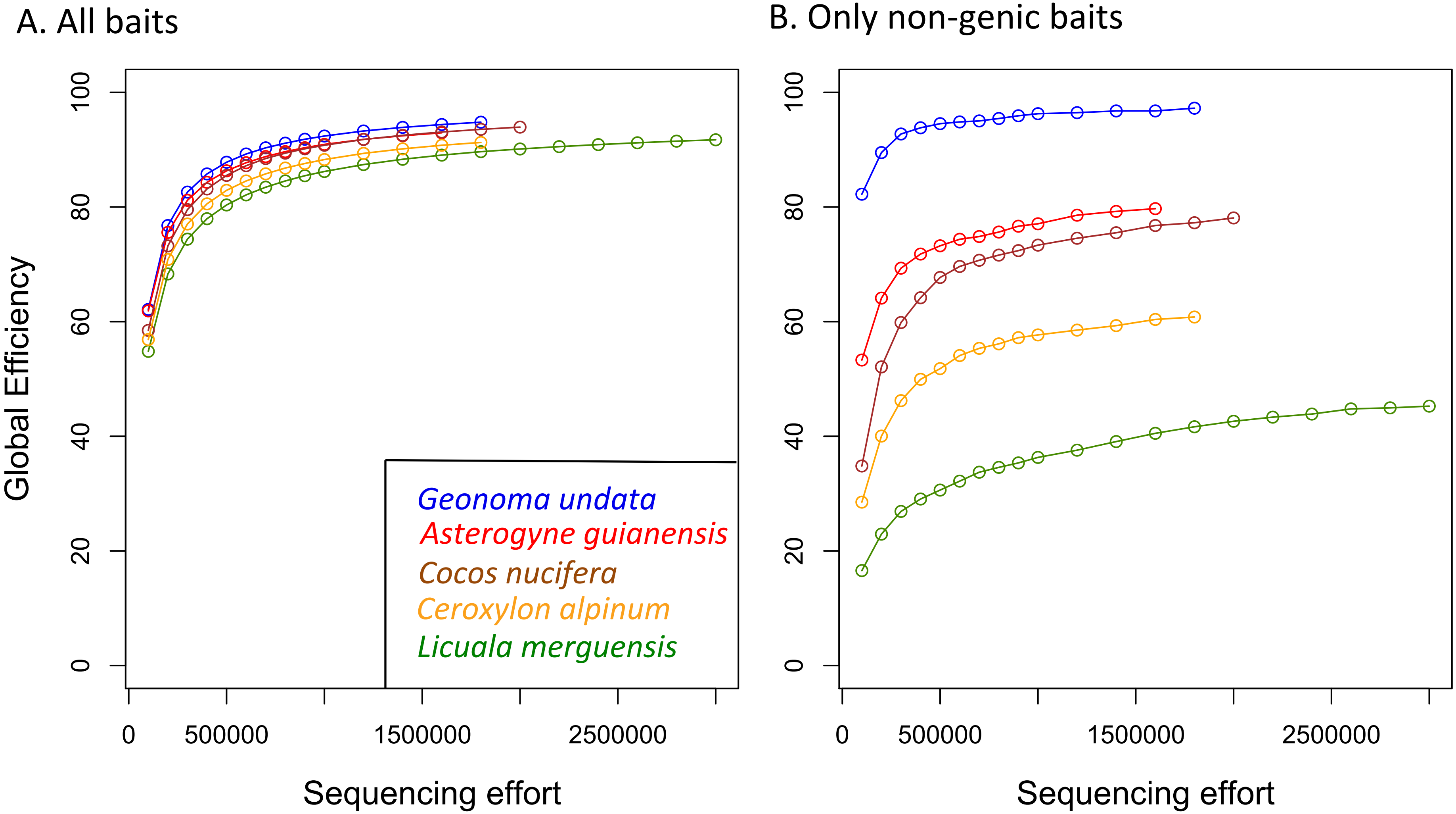
High reproducibility

The all procedure (library preparation + target capture + sequencing) was done in duplicate for each sample to test for reproducibility



High for all sample, for all 3 subfamilies (correlation coefficient range: 0.94 – 0.98)

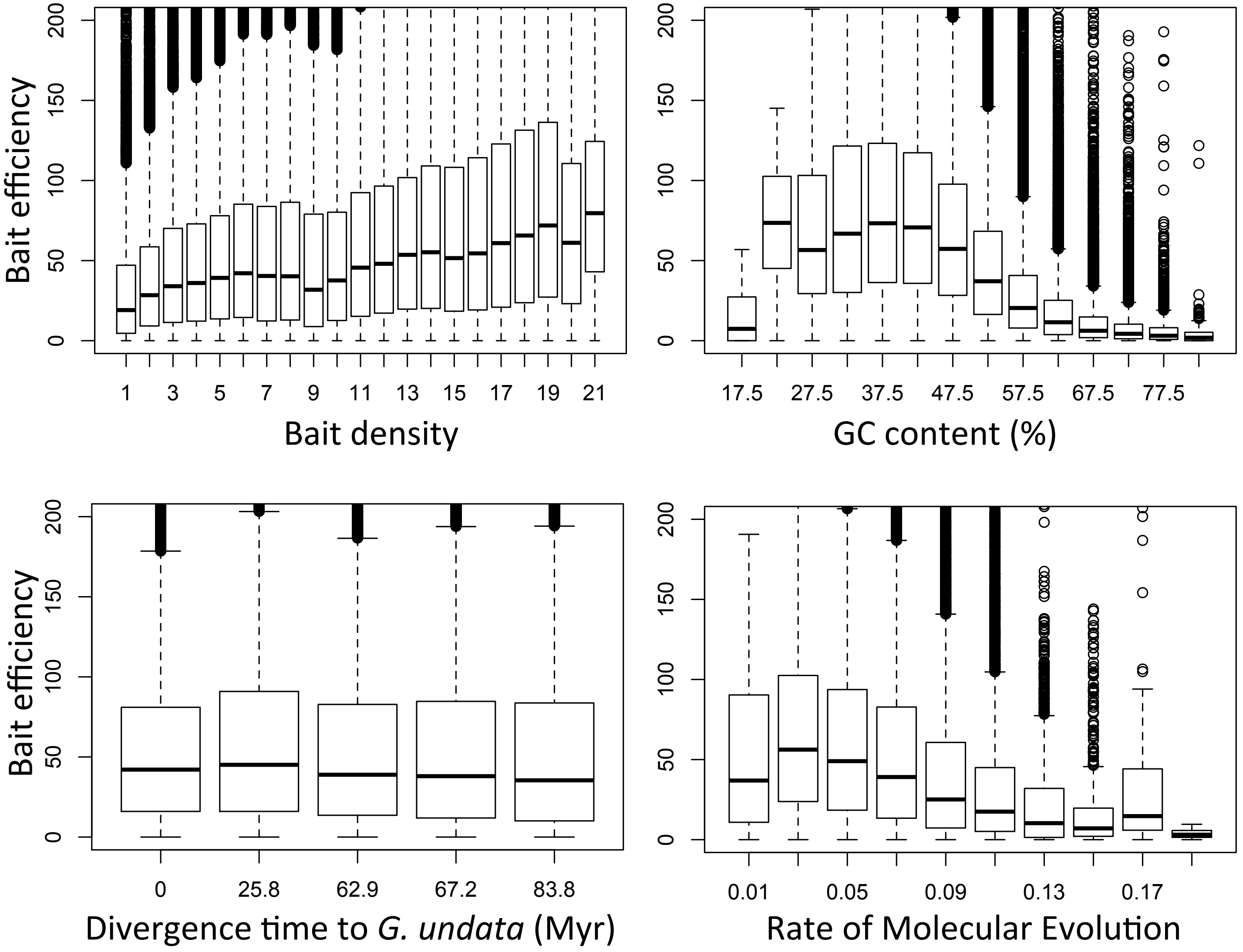
High efficiency of the method



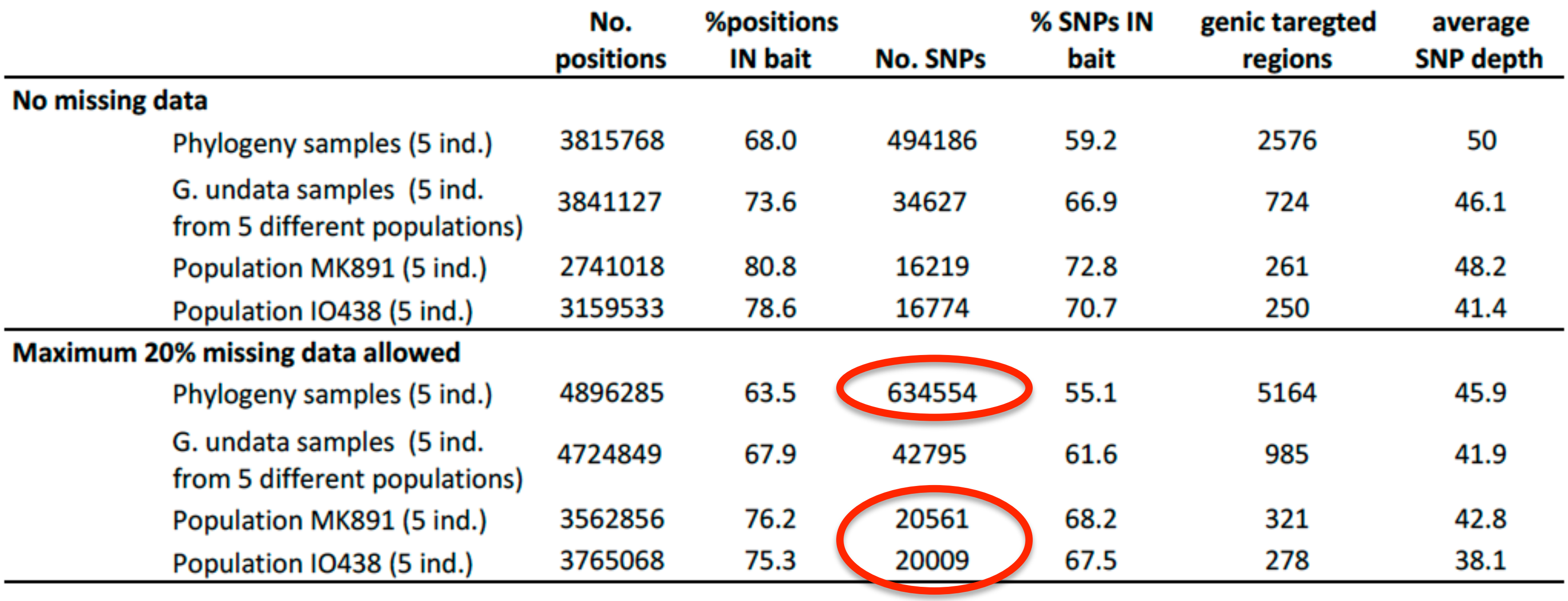
A. All baits

B. Only non-genic baits

Factors influencing bait efficiency



SNP detection

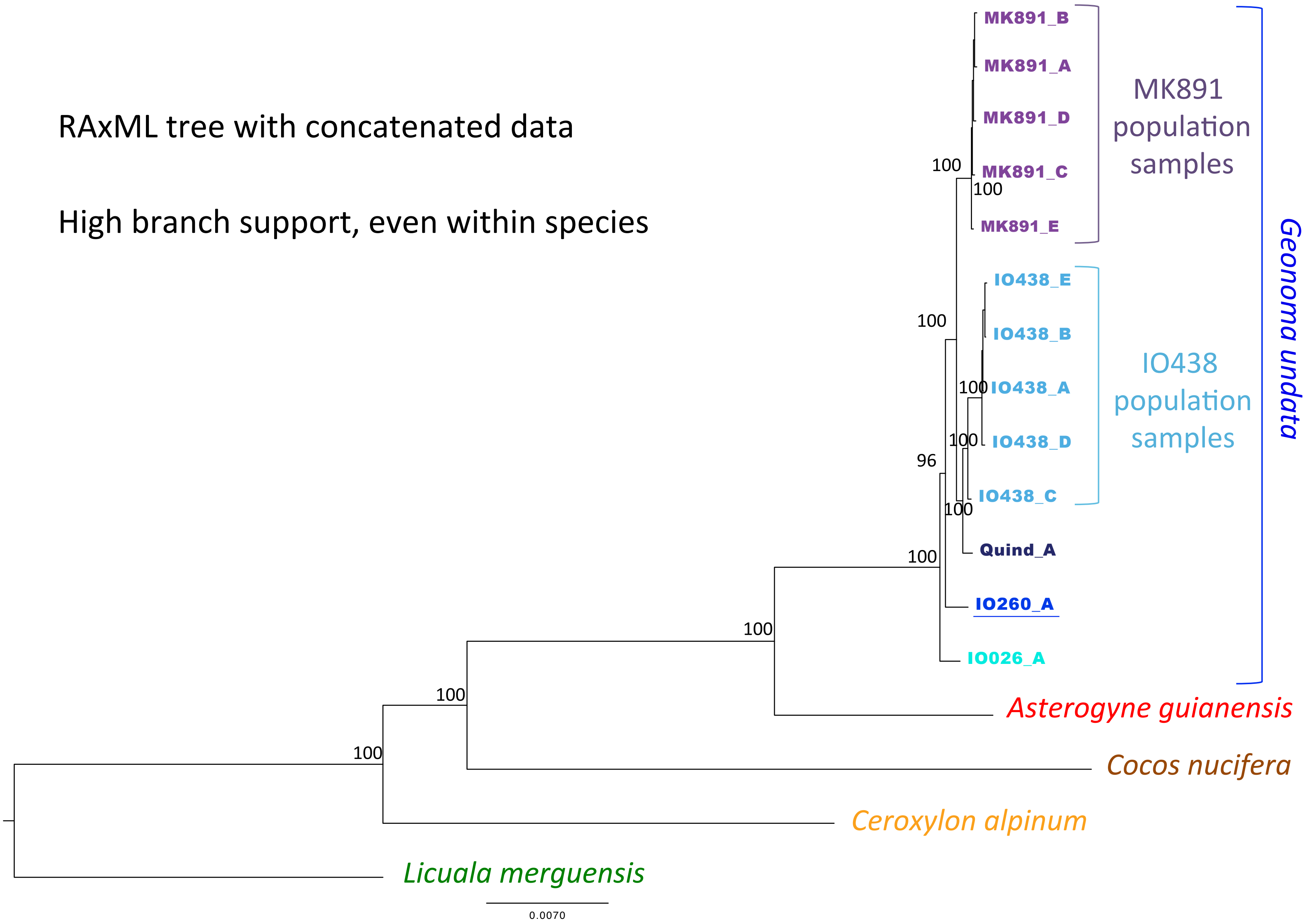


	No. positions	%positions IN bait	No. SNPs	% SNPs IN bait	genic taregted regions	average SNP depth
No missing data						
Phylogeny samples (5 ind.)	3815768	68.0	494186	59.2	2576	50
G. undata samples (5 ind. from 5 different populations)	3841127	73.6	34627	66.9	724	46.1
Population MK891 (5 ind.)	2741018	80.8	16219	72.8	261	48.2
Population IO438 (5 ind.)	3159533	78.6	16774	70.7	250	41.4
Maximum 20% missing data allowed						
Phylogeny samples (5 ind.)	4896285	63.5	634554	55.1	5164	45.9
G. undata samples (5 ind. from 5 different populations)	4724849	67.9	42795	61.6	985	41.9
Population MK891 (5 ind.)	3562856	76.2	20561	68.2	321	42.8
Population IO438 (5 ind.)	3765068	75.3	20009	67.5	278	38.1

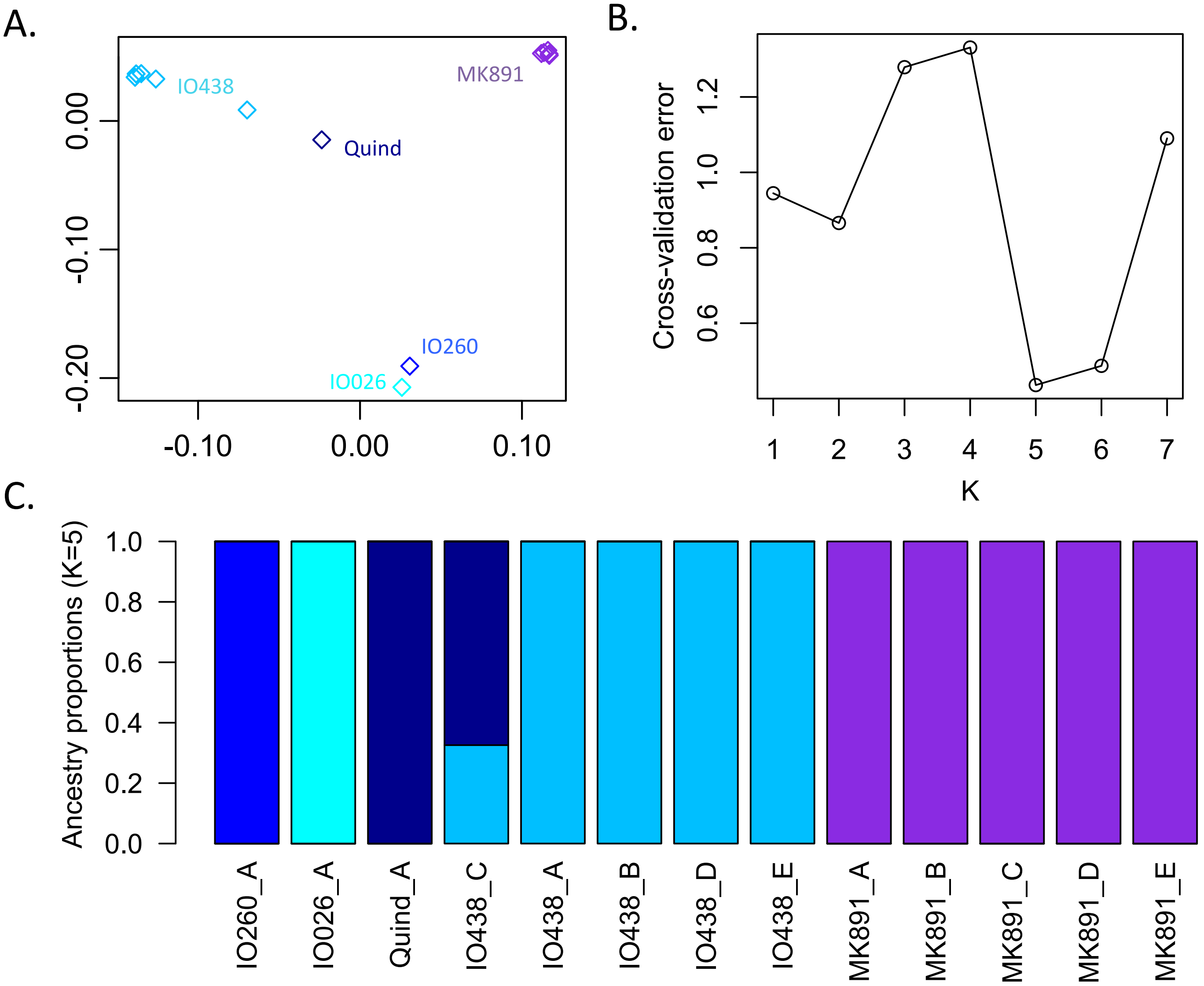
Efficiency of the bait set for phylogeny

RAxML tree with concatenated data

High branch support, even within species



Efficiency of the bait set for population genomics



Ongoing work

Different sets of bait lists :

- Full 60’000 baits = popcorn kit
- 57’000 baits = combine popcorn kit + Heyduck bait set (2015)

Mybait kit 3

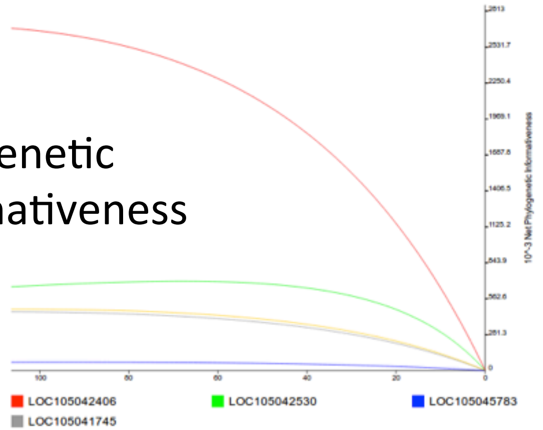
- 57’000 baits = popcorn kit
- 54’000 baits = combine popcorn kit + Heyduck bait set (2015)

Other companies size kit

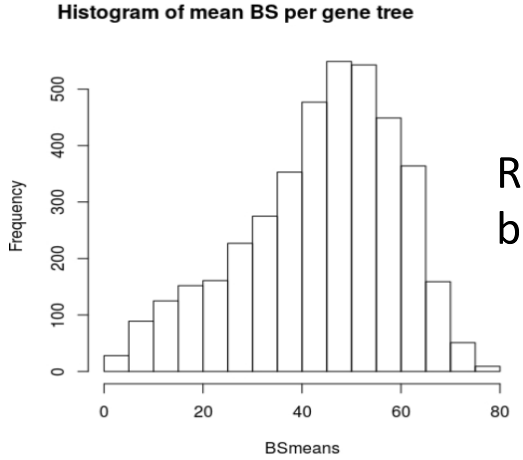
- 20’000 baits = reduced phylogeny informative kit

Mybait kit 1

Phylogenetic informativeness



PhyDesign



RAxML trees, branch support

(Heyduck et al. 2015)

Thanks for your attention

UNI FR

UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

FNSNF
FONDS NATIONAL SUISSE

universität wien

Christian Lexer

Marylaure de la Harpe

POPCORN group